*Environmental Health Perspectives*
*Vol. 61, pp. 229-238, 1985*

# Some Methods of Obtaining Quantitative Structure–Activity Relationships for Quantities of Environmental Interest

## by Marvin Charton*

**Methods are described for obtaining quantitative structure–activity relationships (QSAR) for the estimation of quantities of environmental interest. Toxicities of alkylamines and of alkyl alkanoates are well correlated by the alkyl bioactivity branching equation (ABB). Narcotic activities of 1,1-disubstituted ethylenes are correlated by the intermolecular forces bioactivity (IMF) equation. When the data set has a limited number of substituents in equivalent positions the group number (GN) equation, derivable from the IMF equation, can be used for correlation. It has been successfully applied to aqueous solubilities, 1-octanol–water partition coefficients, and bioaccumulation factors and ecological magnifications for organochlorine compounds. A combination of the omega method for combining data sets for different organisms with the GN equation has been used to correlate toxicities of organochlorine insecticides in two species of fish. Toxicities of carbamates have been correlated by a combination of the zeta method and the IMFB equation. The ABB and the GN equations are particularly useful in that they generally do not require parameter tables, and that the parameters they use are error-free. The methods presented here, as shown by the examples given, should make it possible to establish a collection of QSAR for toxicities, bioaccumulation factors, aqueous solubilities, partition coefficients, and other properties of sets of compounds of environmental interest.**

## Introduction

Two problems of major interest to environmental scientists are the prediction of chemical toxicity and of properties such as bioaccumulation. The most effective method for making estimates of these and similar quantities is correlation analysis. It requires a minimal amount of time, is low cost, and the statistics can easily be obtained using microcomputers and readily available programs. The method is generally applicable to the modeling of the variations of chemical properties and reactivities, physical properties and biological activities with change in molecular structure.

The basis of correlation analysis is the assumption that chemical, physical or biological properties of members of a data set of interest are a linear function of some property of structurally similar members of a reference data set. The reference set is generally used to define a parameter. Values from the data set of interest are then correlated with these parameters by means of simple or multiple linear regression analysis.

*Chemistry Department, School of Liberal Arts and Sciences, Pratt Institute, Brooklyn, NY 11205.

## Electrical Effects

Chemical reactivities and physical properties in solution can usually be completely described in terms of electrical and steric effects. It is convenient to operationally define two classes of electrical effects: 1 localized (field and/or inductive) electrical effects and 2 delocalized (resonance) electrical effects (*1*). They are represented by the $\sigma_I$ and $\sigma_D$ parameters respectively. Four different types of $\sigma_D$ constants are required. We may write the structure of a data set in the form $XGY$ where $X$ is a variable substituent: $Y$ is the active site, an atom or group of atoms at which some quantifiable phenomenon occurs; and $G$ is a skeletal group to which $X$ and $Y$ are bonded. $Y$ and $G$ are usually held constant throughout the data set. The type of $\sigma_D$ constant required to model the delocalized electrical effect in a given system depends on the nature of both G and Y (*1,2*). When both $X$ and $Y$ are bonded to $sp^3$ hybridized C atoms in $G$, no delocalized electrical effect is observed, and the electrical effect is dependent only on $\sigma_D$. When $X$ is bonded to C atoms hybridized $sp^N$ with $1 \leq n \leq 2$ and $Y$ is bonded to a C atom hybridized $sp^3$, the $\sigma^O{}_R$ constants give best results. When both $X$ and $Y$ are bonded to C atoms hybridized $sp^N$, there are three pos-

**Table 1. Dependence of electrical effects on *G* and *Y*.[a]**

| Hybridization, $sp^N$ | | | |
|---|---|---|---|
| $n$,C$^x$ | $n$,C$^y$ | Y | Electrical effect parameters required |
| 3 | 3 | Any | $\sigma_I$ |
| $1 < n < 2$ | 3 | Any | $\sigma_I$, $\sigma_R$ |
| $1 \leq n \leq 2$ | $1 \leq n \leq 2$ | M$^{\delta+}$ | $\sigma_I$, $\sigma_R$ |
| $1 \leq n \leq 2$ | $1 \leq n \leq 2$ | M$^+$ | $\sigma_I$, $\sigma_R^+$ |
| $1 \leq n \leq 2$ | $1 \leq n \leq 2$ | M$^-$ | $\sigma_I$, $\sigma_R^-$ |

[a] C$^x$ is the C atom in *G* to which *X* is bonded, C$^y$ is that to which *Y* is bonded; *M* is the atom of Y which is bonded to *G*.

sible cases depending on the electronic requirements of *Y*. If *Y* is a weak electron acceptor such as a carbonyl group, the $\sigma_R$ constants are used; if it is a strong electron acceptor such as a carbenium ion, $CMe_2^+$, the $\sigma_R^+$ constants are required; if it is a strong electron donor such as the dimethylamino group, the $\sigma_R^-$ constants give best results. The dependence of the electrical effect on the nature of *G* and *Y* is briefly summarized in Table 1. A collection of values of electrical effect substituent constants is available (*1*).

Frequently composite electrical effect parameters are used. Any composite constant $\sigma_T$ may be written as the sum of contributions from the localized and delocalized electrical effects. Thus,

$$\sigma_{TX} = \lambda\sigma_{IX} + \delta\sigma_{DX} + h \quad (1)$$

Values of $\sigma_T$ are characterized by their composition which is conveniently described by the quantity $P_D$, defined as

$$P_D = 100\ \delta/(\lambda + \delta) \quad (2)$$

$P_D$ represents the percent of the total electrical effect which is due to the delocalized effect. Commonly used composite electrical effect substituent constants are the Hammett $\sigma_m$ and $\sigma_p$ values for which $P_D = 28$ and 50, respectively. Composite constants designated $\sigma_n$ have been defined from the equation

$$\sigma_{nX} \equiv \sigma_{IX} + [P_D/(100 - P_D)]\sigma_{DX} \quad (3)$$

The subscript $n$ equals $P_D$ and therefore describes the composition of the constants.

## Steric Effects

Steric effects are conveniently described either by parameters bases on van der Waals radii ($r_V$) or by the branching equations (*3–6*). The $\upsilon$ steric parameter is defined by the equation

$$\upsilon_X \equiv r_{VX} - r_{VH} = r_{VX} - 1.20 \quad (4)$$

Substituents whose steric effect is conformationally dependent cannot be characterized by a single steric parameter as the steric effect they exert will depend on the steric requirements of the active site and of the process undergoing study. Groups of the type $MZ^1{}_2Z^2$ and $MZ^1Z^2Z^3$ such as $CMe_2I$ and CHClMe have conformationally dependent steric effects. One way of representing the steric effects of groups of this type is to define several different sets of effective $\upsilon$ values for use with different processes. Sets of $\upsilon'$ and $\upsilon^*$ values have been defined for this purpose (*7,8*). Alternatively, the branching equations can be used. The simple branching (SB) equation is given by the expression

$$Q_{Ak} = \sum_{i=1}^{m} a_i n_i + a_o \quad (5)$$

where $Q$ is the quantity to be correlated and Ak represents an alkyl group; $n_i$ is the number of branches at all atoms in the alkyl group labeled $i$ and is equal to the number of atoms labeled $i + 1$ (Fig. 1); $a_1$ and $a_o$ are coefficients. The SB equation has been applied successfully not only to alkyl groups but to perfluoroalkyl groups (*9*) and to amino acid side chains, assuming that N, O and S atoms exert the same steric effect as do C atoms (*10*). H atoms are considered to have a negligible effect. The SB equation has some major disadvantages. Of particular importance is the assumption that all of the branches at a given atom exert the same steric effect. This is only an approximation, and often not a very good one. The equation is also not capable of dealing with planar $\pi$-bonded groups. The branching equations are designed for use with atoms that are $sp^3$ = hybridized. Cycloalkyl groups can be handled by calculating effective $n_i$ values for them. In so doing one of the major advantages of the branching equations is lost, the fact that for alkyl groups in particular and for tetrahedral atoms in acyclic groups in general $n_i$ are exact error free parameters. The expanded branching (XB) equation, Eq. (6), permits a more accurate representation of steric effects because it distinguishes between the first, second and third branches at an atom. The XB equation takes the form

$$Q_{Ak} = \sum_{i=1}^{m} \sum_{j=1}^{3} a_{ij} n_{ij} + a_{oo} \quad (6)$$

where the subscript $i$ designates the $i$-th atoms in the group and the subscript $j$ designates the branch at the $i$-th atom. Again, $a_{ij}$ and $a_{oo}$ are coefficients; $n_{ij}$ is equal

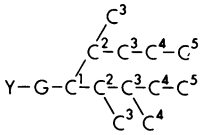


FIGURE 1. An alkyl group numbered for the SB equation. $n_1$ = 2, $n_2$ = 4, $n_3$ = 3, $n_4$ = 2, $n_5$ = 4.

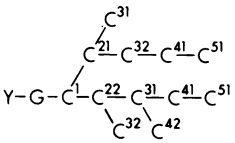


FIGURE 2. An alkyl group numbered for the XB equation. $n_{11}$ = 1, $n_{12}$ = 1, $n_{21}$ = 2, $n_{22}$ = 2, $n_{31}$ = 2, $n_{32}$ = 1, $n_{41}$ = 2, $n_{51}$ = 4.

to the total number of atoms labeled $(i + 1)$ $j$ as is shown in Figure 2. Relative to the SB equation the XB equation has two disadvantages: It requires a much larger data set for good results because of the larger number of independent variables, and, as $n_{11} = 1$ for all alkyl groups except Me, a value of $a_{11}$ generally cannot be determined directly.

Another steric parameter which can sometimes be usefully incorporated into the SB and XB equations is $n_b$, defined as

$$n_b = n_M - 1 \tag{7}$$

where $n_M$ is the number of atoms in the longest chain of the substituent. $n_b$ is a measure of group length.

## Transport Parameters

In a series of papers which represent the most important advance in the quantification of biological activities up to the present time Hansch and his co-workers (*11–14*) found that an equation of the form

$$BA_X = \rho\sigma_X + S\upsilon_X + T_1\tau_X + T_2\tau^2_X + h \tag{8}$$

is generally applicable to biological activities. $\rho$, $S$, $T_1$, $T_2$ and $h$ are coefficients determined by correlating some data set of interest with Eq. (8) by means of multiple linear regression analysis. The $\sigma$ and $\upsilon$ constants have been described above. $\tau$ is a transport parameter. It is convenient to classify transport parameters as primary or secondary. The former include the logarithms of the partition coefficient $P$, the molar solubility (Sol), and the chromatographic flow rate $R_f$. The latter include $\pi$, defined as

$$\pi_X = \log P_X - \log P_H \tag{9}$$

where $P_X$ and $P_H$ are the partition coefficients for the X-substituted and the unsubstituted compounds, respectively; and $R_M$, defined as

$$R_{MX} = [\log (1/R_{fX}) - 1] \tag{10}$$

Transport parameters are a function of the differences in intermolecular forces (imf) between the substrate and phase 1 and those between the substrate and phase 2.

$$\tau = f(\text{imf}_2 - \text{imf}_1) = f(\Delta_{\text{imf}}) \tag{11}$$

The intermolecular forces of interest together with the parameters used to model them are set forth in Table 2. Consider a data set of transport parameters that has a variable substituent $X$. It is possible for $X$ to exert a steric effect on the solvation of functional groups that are in proximity to it. In modeling the transport parameter it is therefore necessary to include a term which

**Table 2. Intermolecular forces.**[a]

| Intermolecular force | Abbreviation | Parameters |
|---|---|---|
| Hydrogen bonding | hb | $n_H$, $n_n$ |
| Dipole-dipole | dd | $\sigma_I$, $\sigma_D$ |
| Dipole-induced dipole | di | $\alpha$, $\sigma_I$, $\sigma_D$ |
| Induced dipole-induced dipole | ii | $\alpha$ |
| Charge transfer | ct | $\sigma_I$, $\sigma_D$ |
| Ion-dipole | Id | $i$ |
| Ion-induced dipole | Ii | $i$ |

[a] $n_H$ and $n_n$ are the number of OH and NH bonds, and the number of full nonbonding orbitals on O or N atoms in the substituent respectively. $i$ takes the value 1 for a charged and 0 for an uncharged group. $\alpha$ is defined by the expression $\alpha_X = (MR_X - MR_H)/100$, where $MR_X$ and $MR_H$ are the molar group refractivities of $X$ and $H$, respectively.

represents the steric effect of $X$. Then using the parameters given in Table 2 and the steric parameter $\upsilon$ we obtain the intermolecular force equation (IMF),

$$Q_X = L\sigma_{IX} + D\sigma_{DX} + A\alpha_X + H_1n_{HX} + H_2n_{nX} + Ii_X + S\upsilon_X + B_o \tag{12}$$

where $Q$ is some transport parameter. In place of the term $S\upsilon_X$ either the SB or the XB equation may be used, giving the alternative relationships

$$Q_X = L\sigma_{IX} + D\sigma_{DX} + A\alpha_X + H_1n_{HX} + H_2n_{nX} + Ii_X + \sum_{i=1}^{m} a_in_i + a_o \tag{13}$$

and

$$Q_X = L\sigma_{IX} + D\sigma_X + A\alpha_X + H_1n_{HX} + H_2n_{nX} + Ii_X + \sum_{i=1}^{m}\sum_{j=1}^{3} a_{ij}n_{ij} + a_{oo} \tag{14}$$

When the substituent $X$ is bonded to an $sp^3$-hybridized C atom, the term in $\sigma_D$ drops out. Transport parameters of amino acids have been successfully correlated with the equation (15):

$$Q_X = L\sigma_{IX} + A\alpha_X + H_1n_{HX} + H_2n_{nX} + Ii_X + S\upsilon_X + B_o \tag{15}$$

Values of log $P$ and $\pi$ for $Ph(CH_2)_nX$, PhX, $XC_6H_4O_2CNHMe$, $XC_5H_4NO_2$, and $XC_5H_4N$ have been correlated with Eq. (12) or relationships derived from it (*15,16*). If X is restricted to alkyl groups, in which case $\sigma_I$ and $\sigma_D$ are constant while $n_H$, $n_n$ amd $i$ are equal to zero, and $\alpha$ is a linear function of the number of carbon atoms in the group ($n_C$), Eqs. (13) and (14) become

$$Q_{Ak} = a_Cn_C + \sum_{i=1}^{m} a_in_i + a_o \tag{16}$$

and

$$Q_{Ak} = a_C n_C + \sum_{i=1}^{m} \sum_{j=1}^{3} a_{ij} n_{ij} + a_{oo} \quad (17)$$

These equations have been applied successfully to over 40 sets of alkyl substituted transport parameters (*18*).

# Results and Discussion

## Alkyl Bioactivity Branching Equation

The Hansch equation [Eq. (8)] may be rewritten in the form

$$Q_X = L\sigma_{IX} + D\sigma_{DX} + Sv_X + T_1\tau_X + T_2\tau_X^2 + h \quad (18)$$

When $X$ is restricted to Ak and $Sv_X$ is replaced by the SB equation we obtain the alkyl bioactivity branching equation (ABB) (*19*):

$$Q_{Ak} = a_C n_C + a_{C^2} n_C^2 + \sum_{i=1}^{m} a_i n_i + a_o \quad (19)$$

As $n_C$ and $n^2_C$ are normally highly collinear, it is useful to rescale $n_C$ thereby breaking the collinearity (*20,21*). We define $n^*_C$ as

$$n_C^* = n_C - (n_{Cmax} + n_{Cmin})/2 \quad (20)$$

where $n_{C,max}$ and $n_{C,min}$ are the maximum and minimum values of $n_C$ in the data set. The ABB equation has been successfully applied to the toxicities of compounds of the type AkY where Y is a group such as —OH, —O—, —OPO(OEt)$_2$, —SPO(Et)$_2$, and —N(NO)— (*22*). Oral $LD_{50}$ values in the rat (mg/kg) for alkylamines were converted to units of mmole/kg and correlated with the ABB equation with $m = 2$. Observed and calculated values are reported in Table 3 (set T51). Best results were obtained on exclusion of the data point for dodecylamine.

**Table 3. Observed and calculated values of data points and their differences: data set T51 for oral $LD_{50}$ (rat) of $AkNH_2$.[a,b]**

| Ak (alkyl group)[c] | Oral $LD_{50}$ (rat), mmole/kg | | |
|---|---|---|---|
| | $Q_o$ | $Q_c$ | $\Delta$ |
| Hx | 0.821 | 0.798 | 0.023 |
| $BuEtCHCH_2$ | 0.542 | 0.696 | –0.054 |
| tBu | 0.391 | 0.455 | –0.064 |
| sBu | 0.716 | 0.647 | 0.060 |
| Bu | 0.835 | 0.839 | –0.004 |
| Am | 0.732 | 0.839 | –0.107 |
| cHx | 0.855 | 0.731 | 0.124 |
| Dc | 0.250 | 0.136 | 0.114 |
| Dd | 1.025 | –0.291 | 1.316 |

[a] $Q_o$ = observed value; $Q_c$ = calculated value; $\Delta = Q_o - Q_c$.
[b] Data of Sax (*23*).
[c] Abbreviations: Hex, hexyl; Am, amyl; Dc, decyl; Dd, dodecyl; tBu, tertiary butyl; sBu, secondary butyl; Bu, *n*-butyl; Et, ethyl; cHx, cyclohexyl.

The observed value seems very high, possibly due to very low water solubility of this compound. $n_2$ is significantly linear in $n_1$ (partial correlation coefficient 0.729, confidence level 95.0%). As $a_2$ was not significant, $n_2$ was excluded from the correlation, giving

$$\log LD_{50,\,Ak} = -0.192(\pm 0.0654)n_1 - 0.101(\pm 0.0200)n_C^* - 0.0201(\pm 0.0111)n_C^{*2} + 0.909(\pm 0.0875)$$

with $F = 10.66$
$s = 0.0993$
$100R^2 = 88.88$
$n = 8$ (21)

The dependence on $n_C^{*2}$ is borderline, probably due to the small size of the data set.

Oral $LD_{50}$ values in the rat for alkyl alkanoates (*23*), again converted to mmole/kg, were correlated with a modification of the ABB equation,

$$Q_{Ak}n = a_C\sum n_C^* + a_{C^2}\sum n_C^{*2} + \sum_{k=1}^{p} \sum_{i=1}^{m} a_{k,i} n_{k,i} + a_o \quad (22)$$

The index $k$ identifies the alkyl group. Thus $n_{2,1}$ refers to the number of branches at $C^1$ in alkyl group 2, for the alkyl alkanoates (I).

$$Ak^1—C(=O)—OAk^2$$

I

As $a_{1,2}$ was not significant $n_{1,2}$ was excluded. The correlation equation obtained is

$$\log LD_{50,Ak^n} = -0.464(\pm 0.140)n_{1,1} - 0.495(\pm 0.162)n_{2,1} + 0.149(\pm 0.0993)n_{2,2} - 0.186(\pm 0.0341)n_C + 3.203(\pm 0.251) \quad (23)$$

**Table 4. Observed and calculated values of data points and their differences: data set T61 for oral $LD_{50}$ (rat) of $Ak^1CO_2Ak^2$.[a,b]**

| $Ak^1$, $Ak^2$ | Oral $LD_{50}$ (rat), mmole/kg | | |
|---|---|---|---|
| | $Q_o$ | $Q_c$ | $\Delta$ |
| Me, Pr[c] | 1.982 | 2.113 | –0.131 |
| Me, iPr | 1.468 | 1.469 | –0.001 |
| Et, Et | 1.535 | 1.500 | 0.035 |
| Me, $EtBuCHCH_2$ | 1.241 | 1.332 | –0.091 |
| Me, Et | 2.097 | 2.150 | –0.053 |
| Me, iBu | 2.111 | 2.076 | 0.035 |
| Me, iHx | 1.630 | 1.555 | 0.075 |
| Pr, Et | 1.276 | 1.314 | –0.038 |
| Me, Bu | 2.083 | 1.927 | 0.156 |

[a] $Q_o$ = observed value; $Q_c$ = calculated value; $\Delta = Q_o - Q_c$.
[b] Data of Sax (*23*).
[c] Abbreviations: iHx, isohexyl; Me, methyl; Et, ethyl; Pr, propyl; iBu, isobutyl.

with $F = 15.55$
$s = 0.125$
$100R^2 = 93.96$
$n = 9$

Observed and calculated values are given in Table 4 (set T61).

## The IMF Equation

The IMF equation [Eq. (12)] and relationships derived from it have been used to correlate log (per cent inhibition) of Gly-Leu hydrolase by amino acids (*24*) and binding of substrates to biopolymers (*16*). Narcotic activities of 1,1-disubstituted ethylenes as measured by the concentration required for 50% of white mice to fall over on one side in a 2-hr exposure were correlated with a relationship derived from the IMF equation,

$$\log BA_{X^1X^2} = L\Sigma\sigma_{IX} + D\Sigma\sigma_{DX} + A\Sigma\alpha_X + H_2\Sigma n_{nX} + S_1\upsilon_{X^1} + S_2\upsilon_{X^2} + B_o \quad (24)$$

As in this data set $L \simeq D$ the composite electrical effect constant $\sigma_{50}$ was used in place of $\sigma_I$ and $\sigma_R$. $S_1$ and $S_2$ were not significant and therefore the terms in $\upsilon_{X^1}$ and $\upsilon_{X^2}$ were dropped. The best correlation equation obtained was

$$\log BA_{X^1X^2} = -1.05(\pm 0.393)\Sigma\sigma_{50,X} + 7.61(\pm 1.51)\Sigma\alpha_X - 0.0987(\pm 0.0314)\Sigma n_{nX} - 2.39(\pm 0.28) \quad (25)$$

with $F = 50.70$
$s = 0.211$
$100R^2 = 94.41$
$n = 13$

Observed (*25*) and calculated values are set forth in Table 5 (set T102).

**Table 5. Observed and calculated values of data points and their differences: data set T102 for concentration for 50% of white mice to fall on one side, $H_2C = CX^1X^2$.[a,b]**

| $X^1X^2$ | Narcotic activities of disubstituted ethylenes | | |
|---|---|---|---|
| | $Q_o$ | $Q_c$ | $\Delta$ |
| Bu,H | -0.807 | -0.796 | -0.011 |
| Am,H | 0.523 | -0.446 | -0.077 |
| Cl,H | -2.284 | -2.241 | -0.043 |
| Cl,Cl | -1.824 | -2.091 | 0.267 |
| OAc,H | -1.250 | -2.075 | 0.815 |
| $O_2CEt$, H | -1.137 | -1.725 | 0.588 |
| $O_2CPr$,H | -0.699 | -1.375 | 0.676 |
| $C_2H$,H | -1.824 | -2.006 | 0.182 |
| $CO_2Me$,H | -1.222 | -2.388 | 1.116 |
| $CO_2Me$,Me | -1.046 | -1.810 | 0.764 |
| CN,H | -2.886 | -2.669 | 0.217 |
| CN,Me | -2.377 | -2.141 | -0.236 |

[a] $Q_o$ = observed value; $Q_c$ = calculated value; $\Delta = Q_o - Q_c$.
[b] Data of Filov et al (*25*).

## The GN Equation

From the IMF equation, assuming that substituents are in equivalent positions and that their effects are additive,

$$Q_{X_m} = mL\sigma_{IX} + mD\sigma_{DX} + mA\alpha_X + mH_1n_{HX} + mH_2n_{nX} + mS\upsilon_X + B_o \quad (26)$$

where $m$ is the number of groups X. For a given substituent X all of the substituent constants are constant. Then, if $m$ is varied

$$Q_{X_m} = (L\sigma_{IX} + D\sigma_{DX} + A\alpha_X + H_1n_{HX} + H_1n_{nX} + S\upsilon_X)m + B_o \quad (27)$$

or

$$Q_{X_m} = B_1m + B_o \quad (28)$$

Then if there are $l$ such $X$ groups in the data set we may write

$$Q_{X_i} = \sum_{i=1}^{l} B_im_{X_i} + B_o \quad (29)$$

If the data set is limited to a few substituents and we assume that the skeletal group can be represented by a term in the number of carbon atoms, $n_C$, we have

$$Q_{X_i} = \sum_{i=1}^{l} B_im_{X_i} + a_Cn_C + B_o \quad (30)$$

which is the group number (GN) form of the IMF equation. The limitation in the number of substituents is determined by the size of the data set as each substituent requires a separate independent variable and the data set must be large enough to provide a sufficient number of degrees of freedom.

A group of compounds of major environmental interest is that which contains C, H, often Cl, and occasionally O. Values of log $P$ and of log $K_B$ (where $K_B$ is the bioaccumulation factor) for these compounds (*26*) were correlated with the GN equation in the form

$$Q_{bas} = B_1m_{Cl} + B_2m'_{Cl} + B_3m_O + a_Cn_c + B_o \quad (31)$$

where $m_{Cl}$ and $m'_{Cl}$ are the number of Cl atoms bonded to $sp^3$ and $sp^2$ hybridized carbon atoms, respectively; $m_O$ is the number of ethereal O atoms. The subscript (bas) indicates bioactive substrate. The correlation equations obtained are

$$\log P_{bas} = 0.227(\pm 0.0161)n_C + 0.250(\pm 0.0230)m_{Cl} + 0.417(\pm 0.0430)m'_{Cl} - 0.447(\pm 0.181)m_o + 1.33(\pm 0.21) \quad (32)$$

with $F = 69.94$
$s = 0.365$
$100R^2 = 90.31$
$n = 35$

and

$$\log K_{B,bas} = 0.212(\pm 0.0164)n_C + 0.198(\pm 0.0234)m_{Cl} + 0.424(\pm 0.0438)m_{Cl}' - 0.496(\pm 0.184)m_O + 0.25(\pm 0.21) \quad (33)$$

with $F = 55.66$
$s = 0.371$
$100R^2 = 88.13$
$n = 35$

In both data sets, $n_C$ and $m_{Cl}'$ are somewhat collinear (partial correlation coefficient = 0.364; confidence level, 95.0%). Though values of $a_C$, $B_2$ and $B_3$ for the two sets are not significantly different the value of $B_1$ is. The two types of Cl atoms were differentiated because the electrical effects of Cl bonded to C will vary with the hybridization state of the C, and the bond moments must therefore vary as well. The range in $P$ is about five orders of magnitude, in $K_B$ it is about four orders of magnitude. Observed and calculated values of log $P$ and log $K_B$ are reported in Table 6 (sets P2 and B1, respectively). Equations (32) and (33) make possible reasonably good estimates of log $P$ and log $K_B$ for a wide range of arenes and organochlorine compounds from the empirical formula of the compound of interest. A very great advantage of both the ABB and the GN equations is that the parameters are error free. A given alkyl group has an exact number of branches at the $i$-th C atoms. A given compound has an exact number of Cl atoms or of O atoms. Furthermore, parameter tables are not required for the use of these equations.

**Table 6. Observed and calculated values of data points and their differences: data set P2 for log $P$ (1-octanol/water) and data set B1 for log $K_B$.[a,b]**

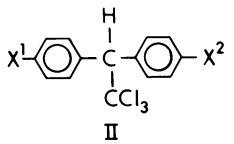
| Biologically active substrate | Log $P$ (octanol/water) | | | Log $K_B$ | | |
|---|---|---|---|---|---|---|
| | $Q_o$ | $Q_c$ | $\Delta$ | $Q_o$ | $Q_c$ | $\Delta$ |
| DDT | 5.75 | 6.09 | –0.34 | 4.47 | 4.66 | –0.19 |
| HC | 5.44 | 5.68 | –0.24 | 4.30 | 4.21 | 0.09 |
| Lindane | 3.85 | 4.19 | –0.34 | 2.67 | 2.71 | –0.04 |
| Mirex | 6.89 | 6.60 | 0.29 | 4.26 | 4.75 | –0.49 |
| CD | 6.00 | 5.93 | 0.07 | 4.58 | 4.41 | 0.17 |
| DDE | 5.69 | 5.84 | –0.15 | 4.71 | 4.46 | 0.25 |
| *o,p*-DDT | 5.75 | 6.09 | –0.34 | 4.57 | 4.86 | –0.09 |
| Dieldrin | 5.48 | 5.44 | 0.04 | 4.11 | 3.94 | 0.17 |
| Endrin | 4.56 | 5.55 | –0.88 | 3.17 | 3.94 | –0.77 |
| HCE | 5.40 | 5.24 | 0.16 | 4.16 | 3.71 | 0.45 |
| PhCl | 3.79 | 3.11 | 0.68 | 2.65 | 1.95 | 0.70 |
| HCBN | 5.28 | 4.84 | 0.44 | 4.05 | 3.35 | 0.70 |
| HCNBD | 5.28 | 4.75 | 0.53 | 3.81 | 3.36 | 0.44 |
| $C_6Cl_6$ | 5.23 | 5.19 | 0.04 | 3.89 | 4.07 | –0.18 |
| $C_2Cl_4$ | 2.88 | 3.45 | –0.57 | 1.59 | 2.37 | –0.78 |
| $Ph_2O$ | 4.21 | 3.61 | 0.60 | 2.29 | 2.30 | –0.01 |
| $CCl_4$ | 2.64 | 2.56 | 0.08 | 1.24 | 1.25 | –0.01 |
| Fluorene | 4.38 | 4.28 | 0.10 | 3.11 | 3.01 | 0.10 |
| PA | 4.46 | 4.51 | –0.05 | 3.42 | 3.22 | 0.20 |
| 2-MePA | 4.86 | 4.73 | 0.13 | 3.48 | 3.68 | –0.20 |
| 124-TCB | 4.23 | 3.94 | 0.29 | 3.32 | 2.79 | 0.53 |
| $CHCl_3$ | 1.95 | 2.31 | –0.36 | 0.78 | 1.06 | –0.28 |
| ACN | 3.92 | 4.05 | –0.13 | 2.59 | 2.79 | –0.20 |
| BaAN | 5.61 | 5.42 | 0.19 | 4.00 | 4.49 | –0.49 |
| 1235-TCB | 4.46 | 4.36 | 0.10 | 3.26 | 3.46 | –0.20 |
| Pyrene | 4.88 | 4.96 | –0.08 | 3.43 | 4.07 | –0.64 |
| 9-MeAN | 5.07 | 4.73 | 0.34 | 3.66 | 3.86 | –0.20 |
| PhH | 2.11 | 2.69 | –0158 | 1.10 | 1.95 | –0.85 |
| AN | 4.34 | 4.51 | –0.17 | 2.96 | 3.64 | –0.68 |
| 4-CDPE | 4.08 | 4.02 | 0.06 | 2.87 | 2.72 | 0.15 |
| 4-CDF | 4.26 | 4.47 | –0.21 | 2.77 | 3.22 | –0.45 |
| $C_6Cl_5H$ | 5.19 | 4.78 | 0.41 | 3.70 | 3.64 | 0.06 |
| NhH | 3.59 | 3.60 | –0.01 | 2.63 | 2.80 | –0.17 |
| $Ph_2$ | 3.88 | 4.05 | –0.17 | 2.64 | 3.04 | –0.40 |
| 14-DCB | 3.53 | 3.53 | 0.00 | 2.33 | 2.37 | –0.04 |

[a] Data of Mackay (*26*) and Metcalf (*27*).
[b] $Q_o$ = observed values: $Q_c$ = calculated values; $\Delta = Q_o - Q_c$.
[c] Abbreviations: HC, heptachlor; CD, chlordane; HCE, heptachlor epoxide; HCNB, heptachloronorbornene; HCNBD, hexachloronorbornadiene; PA, phenanthrene; 124-TCBB, 1,2,4-trichlorobenzene; ACN, acenaphthene; BaAN, benza(A)anthracene; 1235-TCB, 1,2,3,5-tetrachlorobenzene; 9-MeAN, 9-methylanthracene; AN, anthracene; 4-CDPE, 4-chlorodiphenyl ether; 4-CDP, 4-chlorodiphenyl; NhH, naphthalene; 14-DCB, 1,4-dichlorobenzene; MC, methoxychlor.

Values of the ecological magnification (EM), defined as

$$EM \equiv C_{fish}/C_{aq} \tag{34}$$

where $C_{fish}$ and $C_{aq}$ are the concentrations in fish and in their aquatic environment, for 2,2-bis-(4′-substituted phenyl)-1,1,1-trichloroethanes (II) were correlated with the GN equation in the form



II

$$\log EM_{X^1X^2} = a_C n_C + B_1 m_{Cl} + B_2 m_O + B_3 m_S + B_o \tag{35}$$

The species of fish used was *Gambustia affinis (27)*. Results of correlation with Eq. (35) showed that $n_C$ and $m_{Cl}$ are strongly collinear (partial correlation coefficient = 0.803; 98.0% confidence level). The term in $n_C$ was therefore dropped giving the correlation equation

$$\log EM_{X^1X^2} = 1.40(\pm 0.186)m_{Cl} + 0.637(\pm 0.155)m_O - 0.534(\pm 0.169)m_S + 2.00(\pm 0.23) \tag{36}$$

with $F = 41.20$
$s = 0.276$
$100R^2 = 96.87$
$n = 8$

In this correlation only C, Cl, S and O atoms which were part of X were used to determine $m$. Thus, for example, when $X^1 = X^2 = OMe$ $m_O = n_C = 2$ and $m_{Cl} = m_S = 0$. Observed and calculated values are given in Table 7 (set B2).

Aqueous solubilities of some organochlorine insecticides (*27*) were correlated with the GN equation in the form

(37)

**Table 7. Observed and calculated values of data points and their differences: data set B2, log ecological magnification of 4-X¹C₆H₄CHCCl₃C₆H₄X²X-4.[a,b]**

| Substituents $X^1,X^2$ | $Q_o$ | $Q_c$ | $\Delta$ |
|---|---|---|---|
| Cl,Cl | 4.927 | 4.800 | 0.127 |
| OMe,OMe | 3.189 | 3.274 | –0.085 |
| OEt,OEt | 3.186 | 3.275 | –0.088 |
| Me,Me | 2.146 | 2.000 | 0.146 |
| MeS,MeS | 0.740 | 0.932 | –0.192 |
| Me,Cl | 3.146 | 3.400 | –0.254 |
| Me,OEt | 2.602 | 2.637 | –0.035 |
| MeO,MeS | 2.491 | 2.103 | 0.388 |

[a] Data of Metcalf (*27*).
[b] $Q_o$ = observed values; $Q_c$ = calculated values; $\Delta = Q_o - Q_c$.

where $m_{Cl}$ is the total number of Cl atoms and $m_O$ is the number of ethereal O atoms. Equation (31) was not used because every member of the data set except lindane has $m_{Cl}$ equal to 2. Toxaphene was not included in the correlation because it is a complex mixture rather

$$\log Sol_{bas} = -0.237(\pm 0.187)m_{Cl} + 1.10(\pm 0.382)m_O - 0.484(\pm 0.0934)n_C + 5.18(\pm 1.95) \tag{38}$$

with $F = 14.06$
$s = 0.530$
$100R^2 = 87.55$
$n = 10$

Observed and calculated values of log Sol are reported in Table 8 (data set P1). It should be noted that the dependence on $m_{Cl}$ is doubtful.

## The Zeta Method

We now consider the use of a method which permits combining several data sets into one single large set. To illustrate the method we may examine the reaction of a set of substrates $XGY$ with some constant reagent $R_g$. The reaction conditions including the temperature $T$, the pressure $P$, the solvent $S_v$, and the ionic strength $I_s$ are also held constant. As was noted in the introduction the skeletal group $G$ and the active site $Y$ are held constant throughout a data set. Thus, though the quantity $Q$ is a function of all of these variables

$$Q = f(X, G, Y, R_g, T, P, S_v, I_s) \tag{39}$$

all but X are normally held constant. If we write the most general correlation equation we obtain

$$Q = L\sigma_{IX} + D\sigma_{DX} + Sv_X + g\zeta_G + y\zeta_Y + r\zeta_{Rg} + t\zeta_T + p\zeta_P + s_2\zeta_{Sv} + s_2\zeta I_S + h_0^0 \tag{40}$$

**Table 8. Observed and calculated values of data points and their differences: data set P1 for aqueous solubility (ppm).[a,b]**

| Biologically active substrate | Log solubility $Q_o$ | $Q_c$ | $\Delta$ |
|---|---|---|---|
| DDT | –2.921 | –2.781 | –0.140 |
| DDD | –2.699 | –2.544 | –0.155 |
| Aldrin | –0.959 | –2.050 | 1.091 |
| HC | –1.252 | –1.319 | 0.067 |
| Mirex | 0.863 | 0.854 | 0.009 |
| CD | –2.046 | –1.556 | –0.490 |
| DDE | –2.886 | –2.544 | –0.342 |
| Dieldrin | –0.824 | –0.950 | 0.126 |
| Endrin | –0.824 | –0.950 | 0.126 |
| HCE | –0.456 | –0.219 | –0.237 |

[a] Data of Metcalf (*27*).
[b] $Q_o$ = observed values; $Q_c$ = calculated values; $\Delta = Q_o - Q_c$.

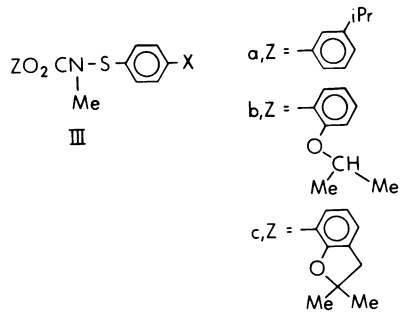
Let us suppose that we wish to combine subsets each of which has a different constant value of one of the variables that represent $G$, $Y$, $R_g$, $T$, $P$, $S_v$, and $I_s$ with all of the other variables held constant. Thus, for example, each subset might have a different value of $G$, though all would have the same values of $Y$, $R_g$, $T$, $P$, $S_v$ and $I_s$. In the general case when X and the factor Z are varied, the combined set can be correlated with the equation

$$Q_{XZ} = L\sigma_{IX} + S\upsilon_X + z\zeta_Z + h_o' \quad (41)$$

The variable $\zeta_Z$ represents the factor $Z$ that remains constant through a subset but varies in the overall set. If a suitable value for $Z$ is not known it can be parameterized internally. One $X$ group for which data points are extant in each subset is arbitrarily chosen as the reference group and denoted $X^*$. Then we define

$$\zeta_Z = Q_X{}^*{}_Z \quad (42)$$

The method is applicable only if all of the subsets are undergoing the same quantifiable phenomenon at the active site $Y$ by the same mechanism. It is useful for generating predictive equations but if internal parameterization is used it cannot explain the dependence of $Q$ on $Z$.




Values of oral $LD_{50}$ in the mouse for the carbamates III(a,b,c) were converted from milligrams per kilogram to millimole per kilogram and combined into a single data set which was correlated with the relationship

$$\log LD_{50,XZ} = L\sigma_{IX} + D\sigma_{DX} + A\alpha_X + H_2 n_{nX} + S\upsilon_X + z\zeta_Z + B_o \quad (43)$$

derived from the IMF equation. Poor results were obtained. The $\zeta_Z$ were defined from $Q_H$. Separate correlations of combinations of subsets IIIa and IIIb, and of IIIb and IIIc with Eq. (43) were then carried out. The IIIa + IIIb combination gave poor results with Z barely significant. The IIIb + IIIc combination gave very good results with $Z$ significant at the 99.9% confidence level though $D$, $A$ and $H_2$ were not significant. Excluding the variables $\sigma_D$, $\alpha$ and $n_n$ gave the correlation equation

$$\log LD_{50,XZ} = -0.359(\pm 0.172)\sigma_{IX} + 0.254(\pm 0.102)\upsilon_X + 0.913(\pm 0.0922)\zeta_X - 0.00433(\pm 0.0932) \quad (44)$$

with $F = 45.78$
$s = 0.130$
$100R^2 = 95.15$
$n = 11$

**Table 9. Observed and calculated values of data points and their differences: data set T401 for oral $LD_{50}$ (mouse) of $ZO_2CNMe\text{-}S\text{-}C_6H_4X\text{-}4$, mmole/kg.[a,b]**

| Z | X | $Q_o$ | $Q_c$ | Δ |
|---|---|---|---|---|
| IIIa | H | –0.236 | –0.211 | –0.025 |
| | Me | –0.623 | –0.075 | –0.548 |
| | tBu | –0.377 | 0.107 | –0.484 |
| | Cl | –0.429 | –0.240 | –0.189 |
| IIIb | H | –0.025 | –0.018 | –0.007 |
| | Me | 0.053 | 0.118 | –0.065 |
| | tBu | 0.369 | 0.300 | 0.069 |
| | Br | –0.054 | –0.022 | –0.032 |
| IIIc | H | –0.047 | –0.860 | –0.087 |
| | Me | –0.485 | –0.725 | –0.170 |
| | tBu | –0.712 | –0.542 | –0.170 |
| | Cl | –0.863 | –0.889 | 0.026 |
| | Br | –0.815 | –0.864 | 0.049 |
| | MeO | –0.983 | –0.877 | –0.106 |
| | CN | –0.975 | –0.963 | –0.012 |

[a] Data of Fukuto (*28*).
[b] $Q_o$ = observed values; $Q_c$ = calculated values; $\Delta = Q_o - Q_c$.

**Table 10. Observed and calculated values of data points and their differences: data set TO1 for log $LC_{50}$ for rainbow trout (*Salmo gairdnerii*) and bluegill (*Lepimus macrochirus*) in ppm for various biologically active substrates.[a,b]**

| | Rainbow trout | | | Bluegill | | |
|---|---|---|---|---|---|---|
| Biologically active substrate | $Q_o$ | $Q_c$ | Δ | $Q_o$ | $Q_c$ | Δ |
| DDT | –1.92 | –1.79 | –0.13 | –1.30 | –1.27 | –0.03 |
| DDD | –2.05 | –2.19 | 0.09 | –1.85 | –1.62 | 0.36 |
| MC | –1.13 | –1.42 | 0.29 | –1.08 | –0.90 | –0.18 |
| Aldrin | –2.21 | –1.96 | –0.25 | –2.00 | –1.44 | –0.56 |
| Dieldrin | –2.51 | –2.44 | –0.07 | –1.82 | –1.92 | 0.10 |
| Endrin | –2.55 | –2.44 | –0.11 | –3.10[d] | –1.92 | –1.18 |
| HC | –1.89 | –2.12 | 0.23 | –1.59 | –1.60 | 0.01 |
| CD | –1.66 | –1.78 | 0.12 | –1.02 | –1.26 | 0.24 |
| Lindane | –1.74 | –1.61 | –0.13 | –1.00 | –1.09 | 0.09 |

[a] Data of Metcalf (*27*).
[b] $Q_o$ = observed values; $Q_c$ = calculated values; $\Delta = Q_o - Q_c$.
[c] Abbreviations: HC, heptachlor; MC, methoxychlor; CD, chlordane.
[d] Excluded from the correlation.

Observed and calculated values of $LD_{50}$ are set forth in Table 9 (set T401). It seems likely that the toxicity mechanism for IIIa is different from that for IIIb and IIIc. Structurally, IIIb and IIIc are very similar to each other and very different from IIIa.

## The Omega Method

Finally, it has been shown that when a set of compounds *XGY* has undergone biological testing in two or more organisms giving two or more data subsets these may be combined into a single data set on the condition that the same mechanism for the biological activity is extant in each subset (*29*). As the biological activity is generally due to the interaction of the substrate with some receptor site on a biopolymer and mutation easily can alter the receptor site the definition of reproducible parameters characteristic of the organism is difficult. It is best to resort to internal parameterization as described above. The method has been applied to $LC_{50}$ toxicity data for organochlorine insecticides in two species of fish, rainbow trout (*Salmo gairdnerii*) and bluegill (*Lepimus macrochirus*) (*27*) using the GN equation in the form

$$Q_{bas} = a_C n_C + B_1 m_{Cl} + B_2 m'_{Cl}{}' + B_3 m_O + O\omega + B_o \quad (45)$$

where $\omega$ is the organism parameter. The latter was defined in this data set as the $Q$ value for heptachlor. Best results were obtained by excluding the value for endrin in bluegill from the correlation. The correlation equation obtained is

$$\log LC_{50,bas} = 0.346(\pm 0.124)m_{Cl} - 0.595(\pm 0.131)m'_{Cl} - 0.480(\pm 0.177)m_0 + 0.256(\pm 0.0735)n_C + 1.73(\pm 0.449)\omega - 1.96(\pm 1.47) \quad (46)$$

with $F = 8.262$
$s = 0.276$
$100R^2 = 78.97$
$n = 17$

Observed and calculated values of log $LC_{50}$ are given in Table 10 (set T01). It should be noted that $n_C$ and $m_{Cl}$ are strongly collinear (partial correlation coefficient = 0.892, 99.9% confidence level).

# Conclusion

The methods presented here provide a means of obtaining quantitative structure property relationships of utility in environmental science and technology. They may be used to predict toxicity, bioaccumulation, aqueous solubility, partition coefficient, and other quantities of interest. The ABB and GN equations are particularly useful in that they do not require (except for cycloalkyl groups in the ABB equation) the use of parameter tables and are free of parameter error. Methods such as those presented here should eventually make possible reasonable estimates of environmental properties of interest for almost any chemical compound.

## REFERENCES


1. Charton, M. Electrical effect substituent constants for correlation analysis. Progr. Phys. Org. Chem. 13: 119–251 (1981).
2. Ehrenson, S., Brownlee, R. T. C., and Taft, R. W. A generalized treatment of substituent effects in the benzene series: a statistical analysis by the dual substituent parameter equation. Progr. Phys. Org. Chem. 10: 1–80 (1973).
3. Charton, M. The upsilon parameter. Definition and determination. Topics Current Chem. 114: 57–91 (1983).
4. Gallo, R. J. Treatment of steric effects. Progr. Phys. Org. Chem. 14: 115–163 (1983).
5. Charton, M. Steric effects. XIII. The composition of the steric parameter as a function of alkyl branching. J. Org. Chem. 43: 3995–4001 (1978).
6. Charton, M. Steric effects on the formation of alkyl radicals and alkyl carbenium ions. J. Chem. Soc. Perkin Trans. II. 1983: 97–104.
7. Charton, M. Steric effects. III. Bimolecular nucleophilic substitution. J. Am. Chem. Soc. 97: 3694–3697 (1975).
8. Charton, M. The nature of the electrical effect of alkyl groups. II. The significance of the Taft $\sigma^*$ values for alkyl groups. J. Org. Chem. 44: 903–906 (1979).
9. Charton, M. Directing and activating effects of triply bonded groups. In: Chemistry of the Functional Groups, Suppl. C (S. Patai, Ed.), John Wiley, New York, 1983, pp. 269–323.
10. Charton, M., and Charton, B. I. The dependence of the Chou-Fasman parameters on amino acid side chain structure. J. Theor. Biol. 102: 121–134 (1983).
11. Hansch, C., Maloney, P. P., Fujita, T., and Muir, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature 194: 178–180 (1962).
12. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., and Streich, M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J. Am. Chem. Soc. 85: 2817–2824 (1963).
13. Hansch, C., and Fujita, T. Rho–sigma–pi analysis. A method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc. 86: 1616–1626 (1964).
14. Hansch, C., and Deutsch, E. W. The use of substituent constants in the study of structure activity relationships in cholinesterase inhibitors. Biochem. Biophys. Acta 126: 117–128 (1966).
15. Charton, M. Volume and bulk parameters. Topics Current Chem. 114: 107–118 (1983).
16. Charton, M. Bulk and steric parameters in bonding and reactivity of bioactive compounds. In: Rational Approaches to the Synthesis of Pesticides. ACS Symp. Ser. 198, American Chemical Society, Washington, DC, pp. 247–278.
17. Charton, M., and Charton, B. I. The structural dependence of amino acid hydrophobicity parameters. J. Theor. Biol. 99: 629–644 (1982).
18. Charton, M. The dependence of transport parameters on structure: Alkyl group variation. In: First International Telesymposium of Medicinal Chemistry. QSAR in Design of Bioactive Compounds, in press.
19. Charton, M. The correlation of biological activities for alkyl substituted systems. In: Proceedings of the Third Congress of the Hungarian Pharmaceutical Society. Budapest, 1979. Akademiai Kiado, Budapest, 1980, pp. 211–220.
20. Berntsson, P. The use of non-correlated log $P$ and $(\log P)^2$ values in quantitative structure activity relationships. Acta Pharm. Suec. 17: 199–208 (1980).
21. Goodford, P. J. Prediction of pharmacological activity by the method of physicochemical-activity relationships. Adv. Pharmacol. Chemother. 11: 51–97 (1973).

22. Charton, M., and Charton, B. I. Quantitative structure activity relationships for environmental toxicity: applications of the alkyl bioactivity branching equation (abstr.). Third Annual Meeting, Society of Environmental Toxicology and Chemistry, Arlington, VA, 1982, p. 47.
23. Sax, N. I. Dangerous Properties of Industrial Materials, 5th Ed. Van Nostrand-Reinhold, New York, 1979.
24. Charton, M., and Charton, B. I. Amino acid properties and bioactivities as a function of side chain structure. In: Quantitative Approaches to Drug Design (J. C. Deardon, Ed.), Elsevier, Amsterdam, 1983, pp. 260–261.
25. Filov, V. A. Golubev, A. A., Liublina, E. I., and Tolokonstev, N. A. quantitative Toxicology. John Wiley, New York, 1979, p. 69.
26. Mackay, D. Correlation of bioconcentration factors. Environ. Sci. Technol. 16: 274–278 (1982).
27. Metcalf, R. L. Organochlorine insecticides: survey and prospects. In: The Future for Insecticides (R. L. Metcalf and J. J. McKelvey, Eds.), Wiley, New York, 1976, pp. 223–285.
28. Fukuto, T. R. Carbamate insecticides. In: The Future for Insecticides (R. L. Metcalf and J. J. McKelvey, Eds.), Wiley, New York, 1976, pp. 313–342.
29. Charton, M. The organism as a variable in QSAR. In: Quantitative Approaches to Drug Design (J. Dearden, Ed.), Elsevier, Amsterdam, 1983, pp. 69–70.